# Computational Inverse Problems with the Quadratic Wasserstein Metric

Kui Ren

Department of Applied Physics and Applied Mathematics

Columbia University

In collaboration with Bjorn Engquist (UT Austin), Yunan Yang (NYU)

July 8, 2020

# Presentation Outline

# Computational Inverse Problems

**Abstract Forward Formulation.** Let us denote by $f$ an operator that maps an unknown function $m : \mathbb{R}^d \to \mathbb{R}$ ($d \geq 1$) to datum $g$, that is

$$f(m) = g.$$

where $f : \mathcal{M} \mapsto \mathcal{G}$ is assumed to be nonlinear in general.

**Inversion Problem Objective.** For a given datum $g^*$, we are interested in finding an $m^*$ such that $f(m^*)$ matches the datum $g^*$, that is, $f(m^*) = g^*$.

# Computational Inverse Problems

**Computational Inversion.** In the absence of analytical inversion formulas, the inverse problem is often solve computationally by taking the $m^*$ that minimizes the discrepancy between the prediction $f(m)$ and the datum $g$, as the solution. That is

$$m^* = \arg\min_m \Phi(m) := \frac{1}{2}\mathfrak{d}^2(f(m), g)$$

for some given metric $\mathfrak{d}$.

**Computational Challenges.** (i) The minimization problem is often non-convex when $f$ is nonlinear; (ii) Computational cost can be high when the evaluation of $f$ is expensive.

# Computational Inverse Problems

**Ideal Metric to Use.** When deciding what metric to use, we consider the following factors: (i) its easiness to compute; (ii) its convexity property; (iii) its sensitivity to high-frequency noise in data $g$; (iv) its sensitivity to $m$ (through $f$).

**Classical $L^2$ Least-Squares.** The $L^2$ metric is the most popular one:

$$\mathfrak{d}(f(m), g) = \|f(m) - g\|_{L^2(\mathbb{R}^d)}$$

due to its mathematical and computational attractions.

# The Quadratic Wasserstein Inversion

**The $W_2$ Distance for Inversion.** The quadratic Wasserstein metric is recently proposed as an alternative to the $L^2$ metric in solving inverse matching problems.

Let $f$ and $g$ be two probability densities on $\mathbb{R}^d$. The square of the quadratic Wasserstein distance between $f$ and $g$, denoted as $W_2^2(f,g)$, is defined as

$$W_2^2(f,g) := \inf_{T \in \mathcal{T}} \int_{\mathbb{R}^d} |\mathbf{x} - T(\mathbf{x})|^2 f(\mathbf{x}) d\mathbf{x},$$

where $\mathcal{T}$ is the set of measure preserving transforms from $f$ to $g$.

# The Quadratic Wasserstein Inversion

The optimal transportation map is determined by the solution of a version of the Monge-Ampère equation.

**Theorem (c.g. Villani 2003)**

*Let $d\mu(\mathbf{x}) = f(\mathbf{x})d\mathbf{x}, d\nu(\mathbf{x}) = g(\mathbf{x})d\mathbf{x}$. The squared Wasserstein metric is given by*

$$W_2^2(\mu,\nu) = \int_X |\mathbf{x} - \nabla u(\mathbf{x})|^2 f(\mathbf{x})d\mathbf{x}$$

*where $u$ is the solution of*

$$\begin{cases} det(D^2 u(\mathbf{x})) = \frac{f(\mathbf{x})}{g(\nabla u(\mathbf{x}))} \ \mathbf{x} \in X \\ u \text{ is convex} \end{cases}$$

# The Quadratic Wasserstein Inversion

**The $W_2$ Distance for Inversion.** The minimization problem is now:

$$m^* = \arg\min_m \Phi(m) := \frac{1}{2} W_2^2(f(m), g).$$

**The Computational Cost.** An immediate observation is that computationally minimizing $W_2^2(f(m), g)$ is much more expensive than minimizing $\|f(m) - g\|_{L^2}^2$ since the evaluation of $W_2^2(f(m), g)$ is very expensive!

# $W_2$ Inversion of a Diffusion Problem

- Take the following diffusion model with absorption coefficient $\sigma$:

$$\begin{aligned} -\Delta u(\mathbf{x}) + \sigma(\mathbf{x})u(\mathbf{x}) &= 0, && \text{in } \Omega \\ \mathbf{n}\cdot\nabla u + u(\mathbf{x}) &= q(\mathbf{x}), && \text{on } \partial\Omega \end{aligned}$$

- Assume that we could measure datum:

$$f(\sigma) := \sigma\, u[\sigma]$$

  where $u[\sigma]$ is used to make it clear that $u$ depend on $\sigma$.

- The inverse problem aims at reconstructing $\sigma$ from measured $f(\sigma)$.

# $W_2$ Inversion: Insensitivity to Noise

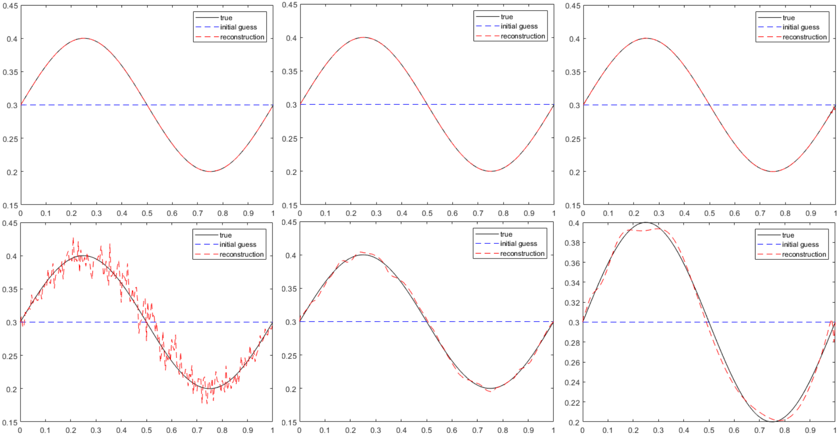


Figure 1: Inversion for $\sigma$ in $\Omega = (0, 1)$ with noise-free (top row) and noisy data (bottom row) under the $L^2$ (left), $\dot{\mathcal{H}}^{-1}$ (middle) and $W_2$ (right) metrics. The noise level in the bottom row is 12% for each case.

# $W_2$ Inversion: Reduction of Resolution

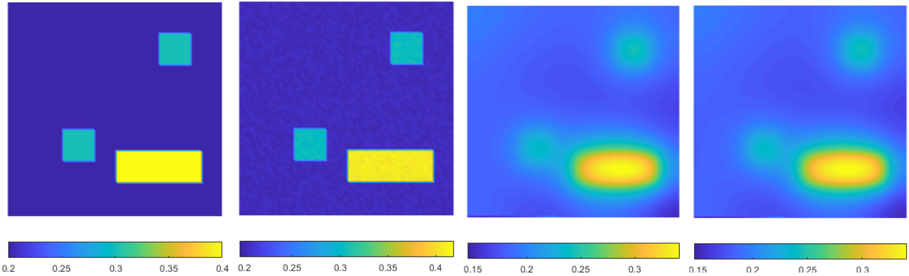


Figure 2: Inversion for $\sigma$ in two-dimensional case in the domain $\Omega = (0,1) \times (0,1)$ with data containing 10% random noise. Shown from left to right are the true coefficient, the reconstructions with the $L^2$, $\mathcal{H}^{-1}$ and $W_2$ metrics respectively.

# Presentation Outline

# The Asymptotic Regime: Background and Notations

**Notations.**

- For any $s \in \mathbb{R}$, let $\mathcal{H}^s(\mathbb{R}^d)$ be the space of functions

$$\mathcal{H}^s(\mathbb{R}^d) := \{m(\mathbf{x}) : \|m\|^2_{\mathcal{H}^s(\mathbb{R}^d)} := \int_{\mathbb{R}^d} \langle \xi \rangle^{2s} |\widehat{m}(\xi)|^2 d\xi < \infty\}.$$

- It is clear that $\mathcal{H}^0(\mathbb{R}^d) = L^2(\mathbb{R}^d)$. When $s \geq 0$, $\mathcal{H}^s(\mathbb{R}^d)$ is the usual Hilbert space of functions with $s$-square integrable derivatives. The space $\mathcal{H}^{-s}(\mathbb{R}^d)$ is understood as the dual of $\mathcal{H}^s(\mathbb{R}^d)$.

# The Asymptotic Regime: Background and Notations

**Notations.**

- We introduce the space $\dot{\mathcal{H}}^1(\mathbb{R}^d)$ with the norm $\|\cdot\|_{\dot{\mathcal{H}}^1(\mathbb{R}^d)}$ defined through the relation

$$\|m\|^2_{\mathcal{H}^1(\mathbb{R}^d)} = \|m\|^2_{L^2(\mathbb{R}^d)} + \|m\|^2_{\dot{\mathcal{H}}^1(\mathbb{R}^d)}.$$

- The space $\dot{\mathcal{H}}^{-1}(\mathbb{R}^d)$ is defined as the dual of $\dot{\mathcal{H}}^1(\mathbb{R}^d)$ via the norm

$$\|m\|_{\dot{\mathcal{H}}^{-1}} := \inf\{|\langle p, m\rangle| : \|p\|_{\dot{\mathcal{H}}^1} \le 1\}.$$

# The Asymptotic Regime: $W_2$ and $\dot{\mathcal{H}}^{-1}$

**Asymptotic Regime.** It was shown [Villani 2003, Section 7.6] that asymptotically $W_2$ is equivalent to $\dot{\mathcal{H}}^{-1}_{(d\mu)}$ (where the subscript $(d\mu)$ indicates that the space is defined with respect to the reference probability measure $d\mu = f(\mathbf{x})d\mathbf{x}$) in the sense that if $\mu$ is a measure and $d\mu$ an infinitesimal perturbation that has zero total mass, then

$$W_2(\mu, \mu + d\pi) = \|d\pi\|_{\dot{\mathcal{H}}^{-1}_{(d\mu)}} + o(d\pi).$$

**Behavior of Global Minimizers.** This result allows us to easily analyze the behavior of $W_2$ inverse solutions (provided that we are lucky enough to find them through minimization).

# Linear Inverse Problems

**Linear Models.** We start with the linear inverse problem give by the model:

$$g = Am\,.$$

When this model is viewed as the linearization of the nonlinear model, $m$ should be understood as the perturbation from the background $m_0$, which is often denoted as $\delta m$.

# Linear Inverse Problems

**Forward Operator.** We assume, without loss of generality, that the linear operator $A$ is diagonal in the Fourier domain and has for symbol (Fourier multiplier)

$$\widehat{A}(\xi) \sim \langle \xi \rangle^{-\alpha}, \quad \langle \xi \rangle := \sqrt{1 + |\xi|^2}.$$

This assumption is not essential at all. It is only made to simplify the calculations.

# Linear Inverse Problems

**Conditioning of Inverse Problem.** Most of inverse problems has exponent $\alpha > 0$. That is, the forward operator $A$ is "smoothing" and therefore the linear inverse problem $Am = g$ is ill-conditioned (so would be the corresponding nonlinear inverse problem $f(m) = g$ when $A$ is thought as the linearization of $f$).

The size of $\alpha$ describes the degree of ill-conditioning of the inverse problem.

The misnomer well-conditioned is often loosely used to refer to the case when $\alpha$ is positive but very small (such as in Radon transform).

# Preconditioning Effects of Weak Norms

**Inverting the Linear Problem with Noisy Data.** Let consider solving the linear inverse problem with noisy data $g^\delta$ in the $\mathcal{H}^s$ least-squares framework.

We invert for $m$ as the minimizer of

$$\begin{aligned} \Phi(m) &:= \frac{1}{2}\|Am - g^\delta\|_{\mathcal{H}^s}^2 \\ &= \frac{1}{2}\int_{\mathbb{R}^d} \langle \boldsymbol{\xi} \rangle^{2s} \left[\widehat{A}\widehat{m} - \widehat{g}^\delta\right]^2 d\boldsymbol{\xi} \end{aligned}$$

# Preconditioning Effects of Weak Norms

**Inverting the Linear Problem with Noisy Data** This is a convex minimization problem whose solution can be written as

$$\widehat{m}(\xi) = \left(\widehat{A}^*(\xi)\big(\langle\xi\rangle^{2s}\widehat{A}\big)\right)^{-1}\widehat{A}^*(\xi)\Big(\langle\xi\rangle^{2s}\widehat{g}^{\delta}(\xi)\Big).$$

where $\widehat{A}^*$ is the $L^2$ adjoint of $\widehat{A}$.

When $s = 0$, the result is simply the classical $L^2$ least-squares solution in the Fourier domain. When $-s > 0$, this is a "preconditioned" version of the $L^2$ solution.

# Preconditioning Effects of Weak Norms

Inversion in Physical Space. The inversion in physical space is simply:

$$m = \Big(A^*(I-\Delta)^s A\Big)^{-1} A^*(I-\Delta)^s g^\delta.$$

where $(I-\Delta)^{s/2}$ is the operator defined through:

$$(I-\Delta)^{s/2} m := \mathcal{F}^{-1}\Big(\langle\xi\rangle^s \widehat{m}\Big).$$

Positive $s$. When $s>0$, $(I-\Delta)^{s/2}$ is a differential operator (nonlocal in the Fourier space). Applying $(I-\Delta)^s$ to the datum $g^\delta$ amplifies high frequency component of the datum.

Negative $s$. When $s<0$, $(I-\Delta)^{s/2}$ is a (smoothing) integral operator. The inversion can be seen as being “preconditioned” in this case.

# Resolution Analysis

Approximate Inverse. Let $R_c$ be an approximation to the inverse of $A$ that is defined through its symbol $\widehat{R}_c(\xi)$:

$$\widehat{R}_c(\xi) = \begin{cases} \langle \xi \rangle^{\alpha}, & |\xi| < \xi_c \\ 0 & |\xi| > \xi_c \end{cases}$$

Reconstruction Error. Let us define $m_0 := R_c g$ as an approximate solution. Following classical results, it is straightforward to verify that

$$\|m - m_0\|_{L^2} \le \|R_c\|_{\mathcal{L}(\mathcal{H}^s;L^2)}\delta + \|(R_c A - I)m\|_{L^2} \le \langle \xi_c \rangle^{\alpha - s}\delta + \langle \xi_c \rangle^{-\beta}\|m\|_{\mathcal{H}^\beta}.$$

under the a priori assumption that $m \in \mathcal{H}^\beta$ for some $\beta > 0$. Here $\delta = \|g^\delta - g\|_{\mathcal{H}^s}$.

# Resolution Analysis

Optimal Error Bound. We can select $\langle \xi_c \rangle = (\delta^{-1} \|m\|_{\mathcal{H}^\beta})^{\frac{1}{\alpha+\beta-s}}$ to minimize the error of the reconstruction, which is bounded by

$$\|m - m_0\|_{L^2} \leq \|m\|_{\mathcal{H}^\beta}^{\frac{\alpha-s}{\alpha+\beta-s}} \delta^{\frac{\beta}{\alpha+\beta-s}}.$$

Optimal Resolution. Therefore reconstruction based on the $\mathcal{H}^s$ framework has a spatial resolution

$$\varepsilon := \langle \xi_c \rangle^{-1} \sim \delta^{\frac{1}{\alpha+\beta-s}}.$$

# Resolution Analysis

Interplay Between Conditioning ($\alpha$), Regularity ($\beta$) and Metric ($s$). The case of $s = 0$ corresponds to the usual reconstruction in the $L^2$ framework. For fixed $\alpha$ and $\beta$, the result says that the resolution of the inversion degenerates when $s$ gets smaller. Therefore, at a fixed noise level, reconstructions in the $\mathcal{H}^s$ framework with $s < 0$ always look smoother.

When $|\alpha|$ or $|\beta|$ (or both) is much larger than $|s|$, the smoothing/desmoothing effect of the metric might not be very visible.

# Resolution Analysis

### Theorem

*Let $R_c$ be an approximation to $A^{-1}$ defined through its symbol:*

$$\widehat{R}_c(\boldsymbol{\xi}) \sim \begin{cases} \langle \boldsymbol{\xi} \rangle^{\alpha}, & |\boldsymbol{\xi}| < \xi_c \\ 0, & |\boldsymbol{\xi}| > \xi_c \end{cases} .$$

*Let $\delta = \|g^{\delta} - g\|_{\mathcal{H}^s}$ be the $\mathcal{H}^s$ norm of the additive noise in $g^{\delta}$. Then the reconstruction error $\|m - m_{\delta}^{c}\|_{L^2}$, with $m_{\delta}^{c} := R_c g^{\delta}$ obtained as the minimizer of $\Phi_{\mathcal{H}^s}(m)$, is bounded by*

$$\|m - m_{\delta}^{c}\|_{L^2} \lesssim \|m\|_{\mathcal{H}^{\beta}}^{\frac{\alpha - s}{\alpha + \beta - s}} \delta^{\frac{\beta}{\alpha + \beta - s}} .$$

*The bound is optimal and is achieved when*

$$\langle \xi_c \rangle^{-1} \sim (\delta \|m\|_{\mathcal{H}^{\beta}}^{-1})^{\frac{1}{\alpha + \beta - s}} .$$

# The Impact of Weighting

**Weighted Norm.** If we weight the $\mathcal{H}^s$ norm with the data to be matched, the minimization problem for the inversion has the objective function:

$$\begin{aligned}\Phi_{H^s_{(d\mu)}}(m) &:= \frac{1}{2}\|Am - g^\delta\|^2_{\mathcal{H}^s_{(d\mu)}}\\ &= \frac{1}{2}\int_{\mathbb{R}^d}\left|\widehat{\sqrt{g^\delta}} \circledast \left[\langle\xi\rangle^s(\widehat{A}\widehat{m} - \widehat{g}^\delta)\right]\right|^2 d\xi.\end{aligned}$$

# The Impact of Weighting

**Inversion with Weighted Norms.** The reconstructed solution in this case, written in physical space, is:

$$m = \Big(A^*(I-\Delta)^{s/2} g^\delta (I-\Delta)^{s/2} A\Big)^{-1} A^*(I-\Delta)^{s/2} g^\delta (I-\Delta)^{s/2} g^\delta.$$

**Impact of Weighting.** The only new thing is the introduction of the inhomogeneity, which depends on the datum $g^\delta$, in the preconditioning operator $(I-\Delta)^s$ (by replacing it with $(I-\Delta)^{s/2} g^\delta (I-\Delta)^{s/2}$).

# Presentation Outline

# The Non-asymptotic Regime

Non-asymptotic Regime. When $f(m)$ is very far away from $g$, that is when initial guess is very far from the true $m$, we do not have a good understanding of what is going on. However, we could still see the main effect of the quadratic Wasserstein metric: the smoothing effect.

The $W_2$ minimization formulation of inverting $f(m) = g$ minimizes:

$$\Phi_{W_2}(m) = \int_{\mathbb{R}^d} |\mathbf{x} - \mathbf{T}(\mathbf{x})|^2 f(m(\mathbf{x})) d\mathbf{x}$$

where $\mathbf{T}$ takes $f(m)$ to $g$.

# The Non-asymptotic Regime

The variation of $\Phi_{W_2}$ with respect to $m$ (along the trajectory of mass conservation) can be found as:

$$\frac{\delta \Phi_{W_2}}{\delta m}(m_k) = \int_{\mathbb{R}^d} \Big( \frac{|\mathbf{x} - \mathbf{T}_k(\mathbf{x})|^2}{2} f'(m_k)[\delta m] \\ - (\mathbf{x} - \mathbf{T}_k(\mathbf{x})) f(\mathbf{x}) \cdot \mathbf{T}'_k \big[ f'(m_k)[\delta m] \big] \Big) d\mathbf{x},$$

where $\mathbf{T}_k$ denotes the optimal transport map at iteration $k$ (that is, for $m_k$), and $\mathbf{T}'_k[\delta f]$ denotes the derivative of $\mathbf{T}_k$ with respect to $f$ (not $m$) in the direction $\delta f$. We have assumed that the map $m \mapsto f$ ($\mathcal{H}^\beta \to \mathcal{C}^{0,\alpha}$) is Fréchet differentiable at any admissible $m$.

# The Non-asymptotic Regime

This gives us that the steepest descent direction of $\Phi_{W_2}$ at $m_k$ is

$$\zeta_k(\mathbf{x}) = f'^*(m_k)\Big[\frac{|\mathbf{x} - \mathbf{T}_k(\mathbf{x})|^2}{2} + \psi(\mathbf{x})\Big],$$

where $f'^*(m_k)$ denotes the $L^2$ adjoint of the operator $f'(m_k)$, and $\psi$ is the weak solution to the (adjoint) linearized Monge-Ampère equation:

$$\sum_{ij} a_{ij}\frac{\partial^2\psi}{\partial x_i \partial x_j} - \sum_j b_j \frac{\partial \psi}{\partial x_j} = -\nabla \cdot \Big(\big(\mathbf{x} - \mathbf{T}_k(\mathbf{x})\big) f(\mathbf{x})\Big).$$

# The Non-asymptotic Regime

Then Caffarelli's regularity theory allows one to show that

**Theorem**

*If $f \in \mathcal{C}^{k,\alpha}(\mathbb{R}^d)$ and $g \in \mathcal{C}^{k,\alpha}(\mathbb{R}^d)$, then*

$$(f(m), g) \mapsto \frac{|\mathbf{x} - \mathbf{T}_k(\mathbf{x})|^2}{2} + \psi(\mathbf{x})$$

*is $\mathcal{C}^{k,\alpha} \mapsto \mathcal{C}^{k+1,\alpha}$.*

Note that the steepest descent direction of $\Phi_{L^2}$ at $m_k$ is

$$\zeta_k(\mathbf{x}) = f'^*(m_k)\Big[f(m_k) - g\Big]$$

# The Non-asymptotic Regime

- In 1D, we can make things explicit. Let $F$ and $G$ be the cumulative density functions for $f$ and $g$ respectively. Then the optimal transportation map from $f$ to $g$ is given by $T(x) = G^{-1} \circ F(x)$.

- This leads to that

$$\zeta_k(\mathbf{x}) = f'^*(m_k)\Big[\frac{(x - T_k(x))^2}{2} - p_k(+\infty) + p_k(x)\Big].$$

with the function $p_k(x)$ is defined as

$$p_k(x) = \int_{-\infty}^{x} \frac{(y - T_k(y))f(m_k(y))}{g(T_k(y))} dy$$

.

# The Non-asymptotic Regime

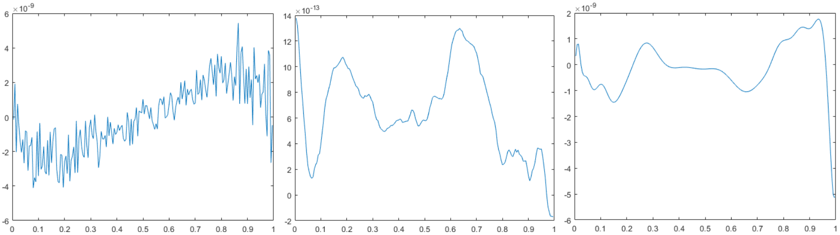

Figure 3: The gradients of the objective functions for the inverse diffusion problem in one-dimensional domain $\Omega = (0, 1)$. Shown are true $m$ (i.e. the absorption coefficient $\sigma$) and two initial guesses (left), the gradients of $\Phi_{L^2}(m)$ at the initial guesses (middle), and the gradients of $\Phi_{W_2}(m)$ at the initial guesses (right).

# The Non-asymptotic Regime

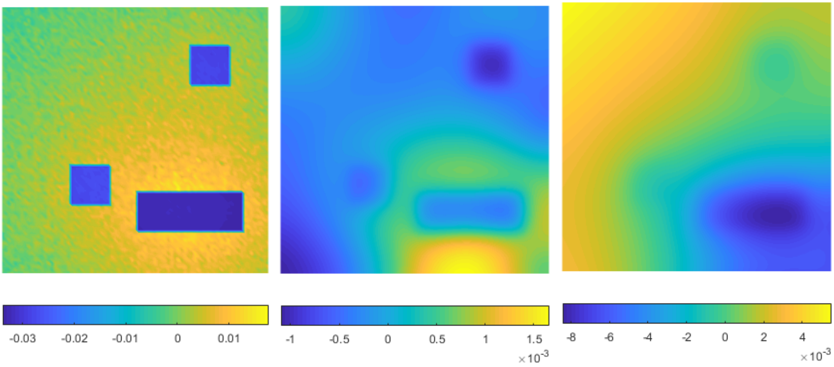


Figure 4: The gradients of the objective functions for the inverse diffusion problem in two-dimensional domain $\Omega = (0,1) \times (0,1)$. Shown are true $m$ (i.e. the absorption coefficient $\sigma$) (top left), the initial guess (top right), the gradient of $\Phi_{L^2}(m)$ at the initial guesses (bottom left), and the gradient of $\Phi_{W_2}(m)$ at the initial guess (bottom right).

# Presentation Outline

# The Quest for Convexity

**$W_2$ Convexifies Some Inverse Problems.** We show here that $W_2$ convexifies certain inverse problems.

**A Simple Transport Model.** Consider a nonlinear problem defined through the following operator, $m := (\eta, \lambda, \mathbf{v})$:

$$f(m)(\mathbf{x}) := \frac{1}{|\eta|^d}\phi(\frac{\mathbf{x} - \lambda\mathbf{v}}{\eta}),$$

where $\phi$ is a given probability density function, $\mathbf{v}$ is a given constant vector, while $\lambda$ and $\eta \neq 0$ are real constants.

# The Quest for Convexity

**A Simple Transport Model.** When $\eta = 1$, this operator serves as a model of transport of a quantity $\phi$ in a given uniform flow $\mathbf{v}$ for a distance $\lambda$. In other words, $\psi(\lambda, \mathbf{v}) := f(m)$ is the solution to the following transport equation at time $\lambda$:

$$\frac{\partial \psi}{\partial t} + \mathbf{v} \cdot \nabla \psi = 0, \text{ in } \mathbb{R}_+ \times \mathbb{R}^d, \qquad \psi(0, \mathbf{x}) = \phi(\mathbf{x}), \text{ on } \mathbb{R}^d.$$

The parameter $\eta$ models the dilation of the transported signal by a factor $\frac{1}{\eta}$ and the factor $\frac{1}{|\eta|^d}$ make sure that the dilated signal is still a probability density function.

# The Quest for Convexity

**Inverse Transport.** For a given function $\phi$, we are interested in finding the unknown $m$ from the datum $g := \frac{1}{|\eta_g|^d}\phi(\frac{\mathbf{x}-\lambda_g \mathbf{v}_g}{\eta_g})$ by matching the predicted datum $f$ with $g$ under the $W_2$ distance.

**Double Convexity.** It is well-known that $W_2^2(f,g)$ is convex with respect to $\lambda \mathbf{v}$ and $\eta$.

$$W_2^2(f,g) = (\eta-\eta_g)^2\mathbb{M}_2 - 2(\eta-\eta_g)(\lambda_g\mathbf{v}_g - \lambda\mathbf{v})\cdot \mathbf{M}_1 + |\lambda_g\mathbf{v}_g - \lambda\mathbf{v}|^2$$

when $\mathbb{M}_2 := \int_{\mathbb{R}^d} |\mathbf{x}|^2\phi(\mathbf{x})d\mathbf{x}$ and $\mathbf{M}_1 := \int_{\mathbb{R}^d} \mathbf{x}\phi(\mathbf{x})d\mathbf{x}$ are finite.

# The Quest for Convexity: A Nonlinear Deconvolution Problem

**A General Convolution Model.** Take the forward operator as a general convolution operator:

$$f(m(\mathbf{x})) \equiv Am(\mathbf{x}) := \int_{\mathbb{R}^d} K(\mathbf{x}-\mathbf{y})m(\mathbf{y})d\mathbf{y}.$$

This type of operators appear in many areas of applications, such as signal and image processing and optical imaging, where $A$ serves as the model of the point spread function of some given physical systems.

**The Deconvolution Problem.** Deconvolution problem aims at inverting $A$ with a given (possibly noisy) datum $g^\delta$ to find $m$. The case with $m$ being a $\delta$ function is of great importance.

# The Quest for Convexity: A Nonlinear Deconvolution Problem

**Small Inclusions/Localized Sources.** In general, let $0 < \varepsilon \ll 1$ be given. We introduce

$$\mathcal{M}_{\mathbf{y},\varepsilon} = \{m(\mathbf{x}) \geq 0 \mid \exists B_\varepsilon(\mathbf{y}) \subseteq \mathrm{supp}(m) \text{ s.t.} \\ \int_{B_\varepsilon(\mathbf{y})} m(\mathbf{x})d\mathbf{x} \geq (1-\varepsilon^{d+1}) \int_{\mathbb{R}^d} m(\mathbf{x})d\mathbf{x}\}$$

where $B_\varepsilon(\mathbf{y})$ denote the ball of radius $\varepsilon$ centered at $\mathbf{y}$. Functions in $\mathcal{M}_{\mathbf{y},\varepsilon}$ have their total mass concentrated in a ball of radius $\varepsilon$, and are therefore highly localized.

# The Quest for Convexity: A Nonlinear Deconvolution Problem

We summarize what is known on this problem in the following theorem.

### Theorem

*Let f and g be generated with* $m_f$ *and* $m_g$ *respectively. Assume that* $m_f$ *and* $m_g$ *have the same total mass. (i) For any kernel function* $K(\mathbf{x})$ *with finite total mass, if*

$$m_\zeta(\mathbf{x}) = \delta(\mathbf{x} - \bar{\mathbf{x}}_\zeta), \ \ \zeta \in \{f, g\}$$

*then we have that*

$$W_2^2(f, g) = |\bar{\mathbf{x}}_f - \bar{\mathbf{x}}_g|^2.$$

*(ii) For any kernel function* $K(\mathbf{x}) \in \mathcal{C}^2(\mathbb{R}^d)$*, if* $m_\zeta \in \mathcal{C}^2(\mathbb{R}^d) \cap \mathcal{M}_{\bar{\mathbf{x}}_\zeta, \varepsilon}$*, then*

$$W_2^2(f, g) = |\bar{\mathbf{x}}_f - \bar{\mathbf{x}}_g|^2 + \mathcal{O}(\varepsilon^{d+1}).$$

# The Quest for Convexity: A Nonlinear Deconvolution Problem

Observation (i). Deconvolution from datum $g$, with an arbitrary kernel $K$, to recover the location of a point source is a convex problem under the $W_2$ metric. This is a simple but NOT obvious observation. Remember that once the source is parameterized in terms of the location, the convolution problem becomes nonlinear (with respect to the location).

Observation (ii). In fact, the source in part (i) does not need to be truly point source. It just needs to be highly localized.

# The Quest for Convexity: A Nonlinear Deconvolution Problem

**Difference to $\dot{\mathcal{H}}^{-1}$.** Recall that the $\dot{\mathcal{H}}^s$ distance between $f$ and $g$ is given by

$$\|f-g\|_{\dot{\mathcal{H}}^s}^2 = 2\int_{\mathbb{R}^d} |\xi|^{2s} |1-\cos(\xi\cdot(\mathbf{x}_f-\mathbf{x}_g))|\,|\widehat{K}(\xi)|^2 d\xi.$$

For a general kernel function $K$, one can only expect this to be a convex function of $\mathbf{x}_f-\mathbf{x}_g$ when $\mathbf{x}_f$ and $\mathbf{x}_g$ are sufficiently close (in which case a first-order Talyor expansion shows the convexity).

# The Quest for Convexity: A Nonlinear Deconvolution Problem

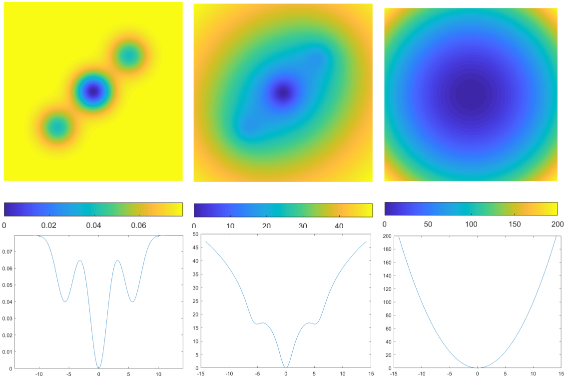


Figure 5: Top row: plots of the $\|f-g\|^2_{L^2}$ (left), $\|f-g\|^2_{\dot{\mathcal{H}}^{-1}}$ (middle) and $W_2^2(f,g)$ (right) for the $m(\mathbf{x})$ given by a Gaussian in the two-dimensional region $\lambda\bar{\mathbf{x}}-\lambda_g\bar{\mathbf{x}}_g \in [-10,10]^2$ with $\Sigma=\Sigma_g=\mathbf{I}_2$. Bottom row: the corresponding cross-sections along the left bottom to top right diagonal.

# Concluding Remarks

There are great interests in recent years on $W_2$ inversion. Since $W_2$ metrics are much more expensive to compute compared to classical metrics we use, "good" properties (if any) of $W_2$ inversion need to be discovered for it to be a legit alternative. Here are key observations from our study:

- In the ideal world, the $W_2$ functional finds the same global minimizers as the $L^2$ functional.
- The $W_2$ inversion is robust w.r.t. high-frequency noise in the data.
- The $W_2$ inversion eliminates high-frequency components of the unknown (which is not a major issue for ill-posed inverse problems since the forward operators already smooth out high-frequency information in the unknown).

# Concluding Remarks

- The $W_2$ functional convexifies certain inverse problems.
- While the smoothing effect of $W_2$ can be replaced by metrics induced by weaker Sobolev norms, it is not clear what metrics share the convexity property that $W_2$ has on some inverse problems.
- Similar effects are also observed in other forms of Wasserstein metrics: $W_1$ compared to $L^1$ (Ding-R. 2020), as well as various modified $W_2$ metrics induced by unbalanced optimal transport (Ding-Du-R. 2020).
- It is of great interests to search for other nonlinear inverse problems that can be convexified with the Wasserstein metrics.